



Helsinki
Center
of
Economic
Research

Discussion Papers

Multiplicative Models for Implied Volatility

Katja Ahoniemi

Helsinki School of Economics, FDPE, and HECER

Discussion Paper No. 172
July 2007

ISSN 1795-0562

HECER – Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI-00014
University of Helsinki, FINLAND, Tel +358-9-191-28780, Fax +358-9-191-28781,
E-mail info-hecer@helsinki.fi, Internet www.hecer.fi

Multiplicative Models for Implied Volatility*

Abstract

This paper estimates a mixture multiplicative error model for the implied volatilities of call and put options on the Nikkei 225 index. Diagnostics show that the mixture multiplicative model is a good fit to the data. In an out-of-sample of two years, multiplicative models correctly predict the direction of change in implied volatility on close to 70 percent of trading days at best. The multiplicative structure also performs better than ARIMA models when forecasting five trading days ahead. An option trading simulation with Nikkei 225 index options points to the predictive superiority of the multiplicative model over ARIMA models.

JEL Classification: C32, C53, G13.

Keywords: Implied volatility, Forecasting, Option markets.

Katja Ahoniemi

Department of Economics,
Helsinki School of Economics
P.O. Box 1210
FI-00101 Helsinki
FINLAND

e-mail: katja.ahoniemi@hse.fi

*The author wishes to thank Markku Lanne, Pekka Ilmakunnas, and Petri Jylhä for valuable comments. Financial support from the FDPE, the Finnish Foundation for Advancement of Securities Markets, and the Yrjö Jahnsson Foundation is gratefully acknowledged. This paper has been presented at the XII Spring Meeting of Young Economists, Hamburg, Germany.

1 Introduction

Reliable volatility forecasts can greatly benefit professional option traders, market makers who need to price derivatives, and all investors with risk management concerns. Implied volatilities, which can be garnered from option markets, can be particularly useful in such contexts as they are forward-looking measures of the market's expected volatility during the remaining life of an option. A correct view of the direction of change in implied volatility can facilitate entering into profitable positions in option markets, and an expected change in the level of market volatility may lead to a need to change stock portfolio weights or composition.

The bulk of implied volatility research investigates how well implied volatility forecasts the volatility that is realized in the returns of an option's underlying asset during the remaining life of the option.¹ Implied volatility has traditionally been modeled with ARMA and linear regression models (e.g. Harvey & Whaley (1992) and Brooks & Oozeer (2002)), or with ARMA models with exogenous regressors and GARCH errors (Ahoniemi (2006)). However, a new class of models, so-called multiplicative models, have been used successfully in recent years to model volatility.

Engle & Gallo (2006), using data on the S&P 500 index, estimate a system of multiplicative error models for squared log returns, the square of the high-low price range, and realized volatility. They conclude that forecasts from the multiplicative specification have significant explanatory power in modeling the value of the VIX index. Lanne (2006) builds a mixture multiplicative error model for the realized volatility of the Deutsche Mark and Japanese Yen against the U.S. dollar. He finds that the in-sample fit of the model is superior to that of ARFIMA models, and forecasts outperform those from several competing models, including ARFIMA and GARCH models.

Multiplicative models are similar in structure to autoregressive conditional duration (ACD) models, which were introduced by Engle & Russell (1998) and have since led to an abundance of research.² So far, multiplicative modeling has not been applied to implied volatility.

This paper models the implied volatility (IV) time series of call and put options on the Nikkei 225 index with a mixture multiplicative model similar to that in Lanne (2006). The model allows for two mean equations and two error distributions, allowing days of large shocks to be modeled separately from more average trading days. The model specification is a good fit to both the call and put IV time series, and produces forecasts with directional accuracy of up to 69.1% in a two-year out-of-sample period. For the Nikkei call IV time series, multiplicative models outperform ARIMA models as directional forecasters, but for the put IV, ARIMA models fare just as well or even slightly better than multiplicative models.

Further analysis nevertheless points to the choice of a multiplicative specification for both call and put IV. When forecasting five trading days ahead, multiplicative models outperform ARIMA models. An option trading simulation using market quotes of Nikkei 225 index options reveals that the best out-of-sample returns are yielded when trading based on the forecasts of multiplicative models.

This paper is structured as follows. Section 2 presents the mixture multiplicative

¹Examples include Day & Lewis (1992), Canina & Figlewski (1993), Jorion (1995), Christensen & Prabhala (1998), and Blair et al. (2001).

²See e.g. Bauwens & Giot (2003), Ghysels et al. (2004), Manganello (2005), Fernandes & Grammig (2006) and Meitz & Teräsvirta (2006).

error model, and Section 3 describes the data used in the study, the estimation results, and diagnostics. Section 4 analyzes the forecasts from various competing models, Section 5 describes the option trading simulation, and Section 6 concludes.

2 The mixture-MEM model

Multiplicative error models (MEM) were first suggested by Engle (2002) for modeling financial time series. Due to the way they are set up, multiplicative models can be used for time series that always receive non-negative values, such as the time interval between trades, the bid-ask spread, trading volume, or volatility. In traditional regression models, logarithms are normally taken from time series data in order to avoid negative forecasts, but this is not necessary with MEM models.

MEM models differ from traditional, linear regression models in that the mean equation μ_t is multiplied with the error term ε_t :

$$y_t = \mu_t \varepsilon_t \quad (1)$$

where IV_t is the implied volatility time series under analysis. Shocks can be assumed to be independent and identically distributed (*iid*) with mean unity from a non-negative distribution. In this particular study, a mixture multiplicative error model (MMEM) similar to that in Lanne (2006) is estimated. In such a specification, there are two possible mean equations:

$$\mu_{1t} = \omega_1 + \sum_{i=1}^{q_1} \alpha_{1i} IV_{t-i} + \sum_{j=1}^{p_1} \beta_{1j} \mu_{1,t-j} \quad (2)$$

$$\mu_{2t} = \omega_2 + \sum_{i=1}^{q_2} \alpha_{2i} IV_{t-i} + \sum_{j=1}^{p_2} \beta_{2j} \mu_{2,t-j} \quad (3)$$

Therefore, μ_t depends on q past observations of implied volatility and p past expected implied volatilities, and the model is denoted MMEM($p_1, q_1; p_2, q_2$). This autoregressive form for the mean equations can help to capture possible clustering in the data. Clustering is often present in financial time series, meaning that high (or low) volatility can persist. The mixture specification is also extended into the error term, with the error terms assumed to come from two gamma distributions with possibly different shape and scale parameters. In other words, μ_{1t} is paired with ε_{1t} , and μ_{2t} is paired with ε_{2t} . Engle (2002) suggested the exponential distribution for the error term, but the gamma distribution is more general, as it nests e.g. the exponential distribution and the χ^2 distribution.

The time-varying conditional mean and possibility for a mixture of two gamma distributions bring considerable flexibility into the model. These elements can help model the fact that in financial time series, periods of business-as-usual alternate with periods of large shocks, which can be captured by the second regime of the model. The probability parameter π ($0 < \pi < 1$) dictates which state the model is in, i.e. the conditional mean is μ_{1t} and errors ε_{1t} with probability π , and the conditional mean is μ_{2t} and errors

ε_{2t} with probability $(1 - \pi)$. There could of course be even more mixture components in the model than just two, but the diagnostics presented in Section 3 indicate that two components is sufficient for this application.

The conditional mean equations reveal that MEM (and ACD) models are similar in structure to GARCH models, meaning that parameter constraints that apply to GARCH models also apply to MEM models. The estimated coefficients must therefore satisfy the positivity constraints outlined in Nelson & Cao (1992) for GARCH models. For both mixture components of the MMEM(1,2;1,2) model, the constraints are:

$$\omega_i \geq 0$$

$$\alpha_{i1} \geq 0$$

$$0 \leq \beta_i < 1$$

$$\beta_1 \alpha_{i1} + \alpha_{i2} \geq 0$$

with $i = 1, 2$. Therefore, in contrast to a (1,1) model, not all parameters need to be non-negative.

The shape and scale parameters of the gamma distributions are constrained so that with $\varepsilon_{1t} \sim \text{Gamma}(\gamma_1, \delta_1)$, $\delta_1 = 1/\gamma_1$ and with $\varepsilon_{2t} \sim \text{Gamma}(\gamma_2, \delta_2)$, $\delta_2 = 1/\gamma_2$, or so that the scale parameter is the inverse of the shape parameter, which ensures that the error term will have mean unity. Under these assumptions, the conditional distribution of IV_t is:

$$\begin{aligned} f_{t-1}(IV_t; \theta) = & \pi \frac{1}{\mu_{1t} \Gamma(\gamma_1) \delta_1^{\gamma_1}} \left(\frac{IV_t}{\mu_{1t}} \right)^{\gamma_1-1} \exp\left(-\frac{IV_t}{\delta_1 \mu_{1t}} \right) + \\ & (1 - \pi) \frac{1}{\mu_{2t} \Gamma(\gamma_2) \delta_2^{\gamma_2}} \left(\frac{IV_t}{\mu_{2t}} \right)^{\gamma_2-1} \exp\left(-\frac{IV_t}{\delta_2 \mu_{2t}} \right) \end{aligned} \quad (4)$$

where θ is the parameter vector and $\Gamma(\cdot)$ is the gamma function. The model can be estimated with conditional maximum likelihood (ML), and the log-likelihood function can then be written as:

$$\ell(\theta) = \sum_{t=1}^T \ln[f_{t-1}(IV_t)] \quad (5)$$

3 Estimation

3.1 Data

The underlying asset for the option data used in this study is the Nikkei 225 index, which is a price-weighted average of 225 Japanese companies listed on the Tokyo Stock Exchange and likely to be the most closely followed stock index in Asian markets. The

currency of denomination for the Nikkei 225 index is the Japanese Yen. The component stocks of the index are reviewed once a year. The Nikkei 225 reached its all-time high in December 1989, topping 38,900 at the time. In the sample used in this study, the index value ranges from 7,608 to 23,801.

Data on the implied volatility of options on the Nikkei 225 index was obtained from the Bloomberg Professional Service for both Nikkei 225 index call and put options for the time period 1.1.1992 - 31.12.2004. The graphs of the time series of Nikkei 225 call and put IV are shown in Figure 1. The use of separate time series of IV from calls and puts can offer new insights into the analysis, and e.g. benefit investors wishing to trade in only either call or put options.

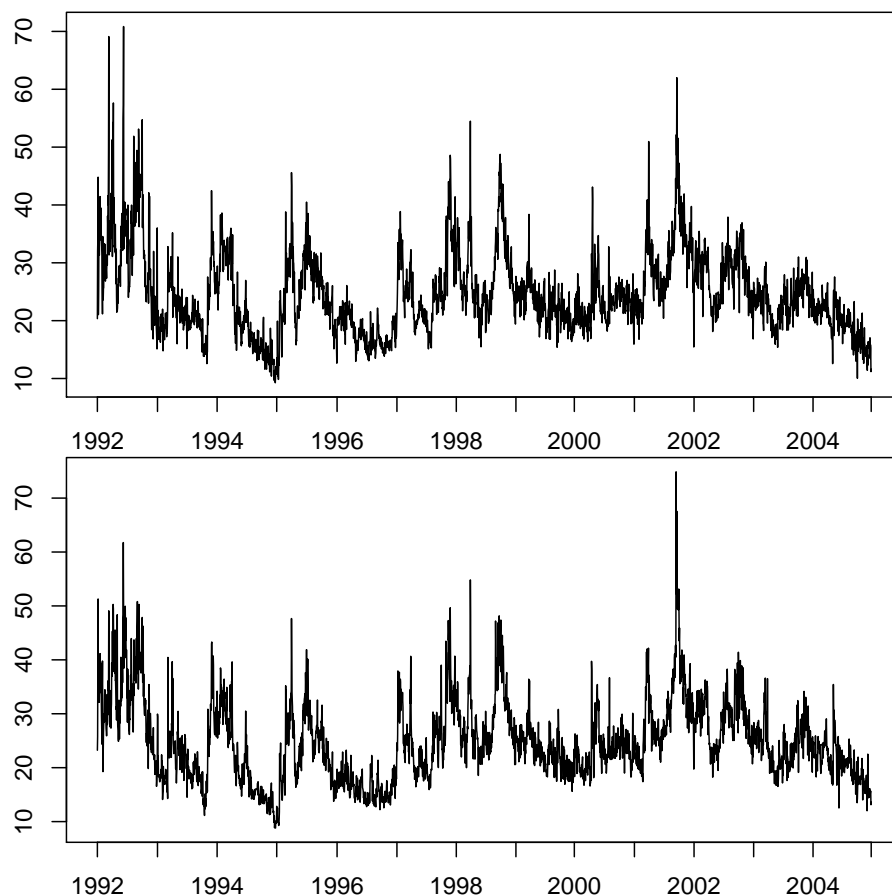


Figure 1: Nikkei 225 index call implied volatility (upper panel) and put implied volatility (lower panel) 1.1.1992 - 31.12.2004.

Although the implied volatility calculated from call and put options with equal expiration dates and strike prices should theoretically be the same, as both reflect the same market expectation, the empirical finding is that the two almost always differ slightly. This may be a reflection of differing demand and supply balances in the markets for calls and puts: for example, Bollen & Whaley (2004) note that in the S&P 500 index option market, puts account for 55% of trades, with institutional investors buying puts as portfolio insurance. They also find that the level of at-the-money implied volatility is largely driven by the demand for at-the-money puts. In the data set used in this study, the time series for put-side IV reacts particularly strongly on 9/11, which is a logical

reflection of the plummet in stock prices and the ensuing panic selling that took place at the time. This high market uncertainty would have raised the demand for put options more than the demand for call options.

The IV time series are calculated daily as the unweighted average of the Black-Scholes implied volatilities of two near-term nearest-to-the-money options. Near-term options tend to be most liquid, and therefore have the most accurate prices. Options on the Nikkei 225 index are available with maturity dates for every month. Days when public holidays fall on weekdays, or when there was no change in the value of call or put implied volatility, were omitted from the data set. After this modification, the full sample contains 3,194 observations.

Descriptive statistics for the Nikkei 225 call (NIKC) and put (NIKP) implied volatility time series are given in Table 1. The IV of puts has been slightly more volatile during the time period in question. Both series are skewed to the right and they display excess kurtosis. The autocorrelations for NIKC and NIKP are displayed in Figure 2, revealing the relatively high degree of persistence in the data. However, a unit root is rejected by the Augmented Dickey-Fuller test for both NIKC and NIKP at the one-percent level of significance. Lanne (2006) finds that with such slowly decaying autocorrelations, a multiplicative model with two mixture components can fit the data quite well.

	<i>NIKC</i>	<i>NIKP</i>
Maximum	70.84	74.87
Minimum	9.26	8.80
Mean	24.68	24.82
Median	23.42	23.84
Standard deviation	7.07	7.41
Skewness	1.10	0.94
Excess kurtosis	2.42	1.79

Table 1: Descriptive statistics for NIKC and NIKP for the full sample of 1.1.1992 - 31.12.2004.

3.2 Model estimation

The in-sample period used in model estimation covers 2,708 observations from 2.1.1992 to 30.12.2002. The base case in the estimation was the MMEM(1,2;1,2) model, which was found to be the best specification for exchange rate realized volatility time series by Lanne (2006). However, the coefficient for α_{22} turned out to not be statistically significant for NIKC or NIKP, so a (1,2;1,1) specification was also estimated for both time series. Table 2 presents the coefficients and log-likelihoods of the estimated models. As the change in log-likelihoods is minimal, it can be concluded that the parameter α_{22} can be dropped from the models.³ The estimated coefficients satisfy all the Nelson & Cao (1992) constraints.

The constant terms in all the estimated models are higher in the second regime, which indicates that this regime most likely models days with larger shocks. The persistence of the model, reflected in the parameters β_1 and β_2 , is almost equal in the first regime for both NIKC and NIKP. NIKC is somewhat more persistent in the second regime. The incidence of the business-as-usual regime is clearly higher for put-side implied volatility

³The p-values from a likelihood ratio test for this restriction are 0.81 for NIKC and 0.53 for NIKP.

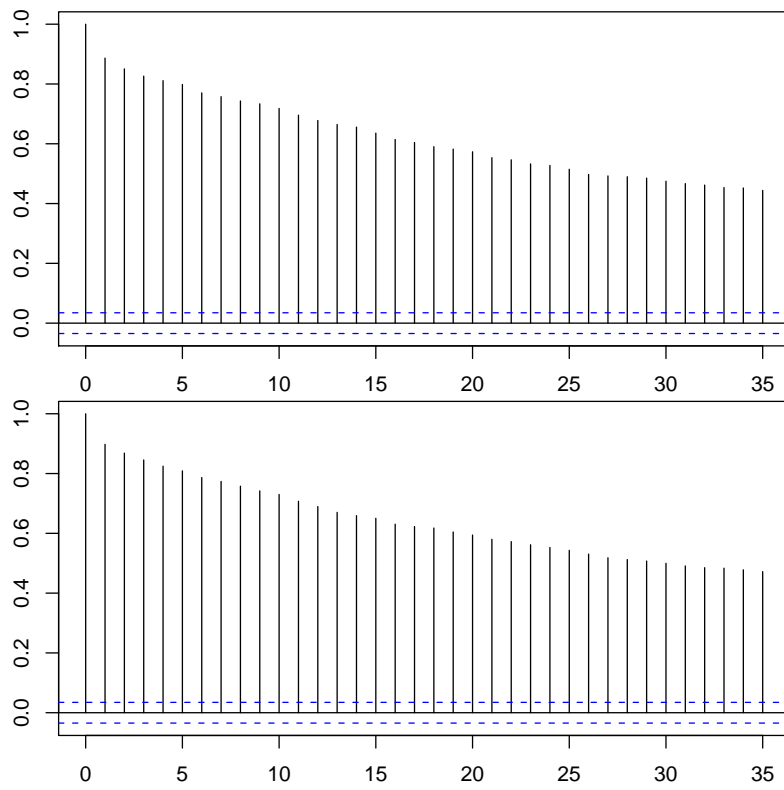


Figure 2: Autocorrelations for NIKC (upper panel) and NIKP (lower panel). The dashed lines mark the 95% confidence interval.

	<i>NIKC</i>		<i>NIKP</i>	
	<i>MMEM(1,2;1,2)</i>	<i>MMEM(1,2;1,1)</i>	<i>MMEM(1,2;1,2)</i>	<i>MMEM(1,2;1,1)</i>
Log likelihood	-6422.82	-6422.85	-6384.90	-6385.10
π	0.821 (0.000)	0.822 (0.000)	0.940 (0.000)	0.940 (0.000)
γ_1	145.689 (0.000)	145.495 (0.000)	113.177 (0.000)	113.300 (0.000)
ω_1	0.264 (0.002)	0.261 (0.002)	0.288 (0.000)	0.292 (0.000)
α_{11}	0.637 (0.000)	0.638 (0.000)	0.573 (0.000)	0.571 (0.000)
α_{12}	-0.255 (0.000)	-0.261 (0.000)	-0.183 (0.000)	-0.177 (0.000)
β_1	0.606 (0.000)	0.610 (0.000)	0.595 (0.000)	0.590 (0.000)
γ_2	26.553 (0.000)	26.541 (0.000)	19.437 (0.000)	19.346 (0.000)
ω_2	0.674 (0.092)	0.717 (0.060)	3.528 (0.240)	2.228 (0.148)
α_{21}	0.339 (0.000)	0.324 (0.000)	0.543 (0.021)	0.587 (0.005)
α_{22}	-0.031 (0.808)	- -	0.324 (0.491)	- -
β_2	0.674 (0.000)	0.657 (0.000)	0.119 (0.810)	0.409 (0.001)

Table 2: Estimation results for the $MMEM(1,2;1,2)$ and $MMEM(1,2;1,1)$ models for *NIKC* and *NIKP*. P-values for the significance of the coefficients are given in parentheses.

than for call-side IV, as the probability parameter receives the value of 0.94 for NIKP and 0.82 for NIKC.

Figure 3 shows the estimated densities of the error terms from the MMEM(1,2;1,1) model for NIKC and NIKP. The densities for the more common, business-as-usual component of the model are more concentrated around unity, and the densities for the second mixture component are more dispersed and skewed to the right. The graphs for the MMEM(1,2;1,2) model specification look essentially the same, as the values for the gamma distribution parameters are nearly identical for the two models (see Table 2).

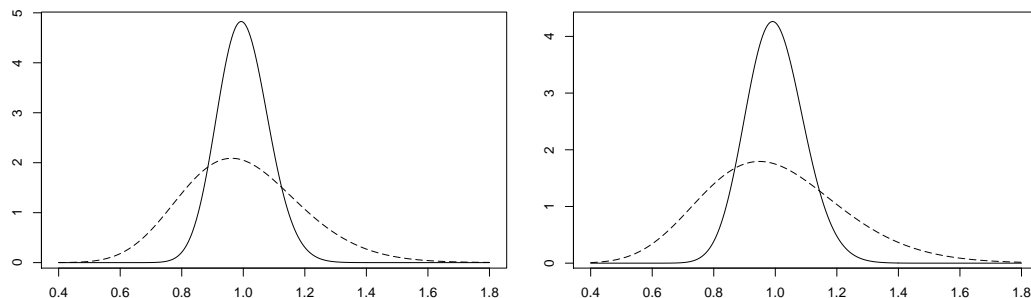


Figure 3: Densities of error terms from the MMEM(1,2;1,1) model for NIKC (left) and NIKP (right). The solid line is the density of ε_{1t} and the dashed line is the density of ε_{2t} .

With financial market data, it may be that only the most recent history is relevant in modeling and forecasting, so the MMEM(1,2;1,2) and MMEM(1,2;1,1) models were also estimated for NIKC and NIKP using only 500 observations from the end of the in-sample period. This corresponds to an in-sample of 20.12.2000 - 30.12.2002. The estimation results for this sample period are given in Table 3. As before, α_{22} is constrained to zero in the models (the p-value for this restriction in an LR test was 0.41 for NIKC and 0.39 for NIKP).

Compared with the values in Table 2, the probability parameters are closer in value between NIKC and NIKP in the models estimated with 500 observations (0.87 and 0.93, respectively). Also, the gamma distribution parameters are now equal in the first regime, contrary to estimates with the entire in-sample. The coefficient for the second lag of NIKP is no longer negative in the first regime, nor is it statistically significant. The constant in the second regime falls to its lower limit, or 0.00001, for NIKC, thus making it lower in value than in the first regime.

In order to investigate the necessity of the mixture components of the model, a MEM(1,2) specification (i.e., a model with only one mean equation and error distribution) was also estimated for both the call and put IV time series. The results of this estimation are given in Table 4. Coefficients are not statistically significant, and the parameters of the gamma distribution are very different from those for the MMEM models (see also Figure 4). This lends support to the model specification with two mixture components. Also, this specification receives further support from likelihood ratio tests for the two mixture components having the same dynamics. The hypothesis of $\omega_1 = \omega_2$, $\alpha_{11} = \alpha_{21}$, $\alpha_{12} = \alpha_{22}$, and $\beta_1 = \beta_2$ is rejected with a p-value of less than 0.0001 for both NIKC and NIKP. Also, the error terms need different distributions in the two regimes, as the hypothesis that $\gamma_1 = \gamma_2$ is rejected with a p-value of less than 0.00001 for both call and put implied volatility.

	<i>NIKC</i>		<i>NIKP</i>	
	<i>MMEM(1,2;1,2)</i>	<i>MMEM(1,2;1,1)</i>	<i>MMEM(1,2;1,2)</i>	<i>MMEM(1,2;1,1)</i>
Log likelihood	-1227.03	-1227.37	-1200.32	-1200.69
π	0.871 (0.000)	0.870 (0.000)	0.928 (0.000)	0.930 (0.000)
γ_1	145.999 (0.000)	146.208 (0.000)	146.099 (0.000)	145.589 (0.000)
ω_1	0.644 (0.011)	0.654 (0.011)	0.894 (0.023)	0.880 (0.026)
α_{11}	0.505 (0.000)	0.491 (0.000)	0.458 (0.000)	0.452 (0.000)
α_{12}	-0.268 (0.000)	-0.248 (0.001)	0.035 (0.653)	0.042 (0.597)
β_1	0.741 (0.000)	0.735 (0.000)	0.475 (0.000)	0.475 (0.000)
γ_2	27.251 (0.000)	27.471 (0.000)	24.841 (0.004)	24.656 (0.004)
ω_2	0.000 (0.000)	0.000 (0.060)	1.495 (0.728)	1.522 (0.667)
α_{21}	0.487 (0.077)	0.633 (0.006)	0.367 (0.367)	0.548 (0.060)
α_{22}	0.469 (0.446)	- -	0.411 (0.456)	- -
β_2	0.025 (0.969)	0.353 (0.139)	0.223 (0.680)	0.427 (0.167)

Table 3: Estimation results for the MMEM(1,2;1,2) and MMEM(1,2;1,1) models for NIKC and NIKP with an in-sample of 500 observations. P-values for the significance of the coefficients are given in parentheses.

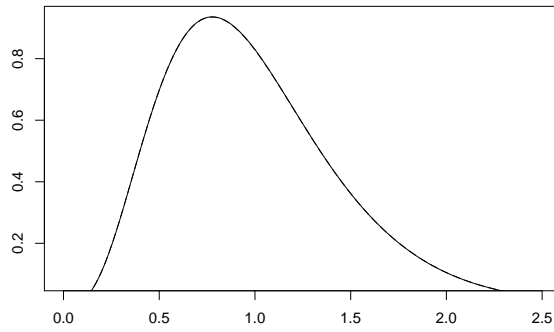


Figure 4: Density of error terms from the MEM(1,2) model for NIKC and NIKP.

	<i>NIKC</i>	<i>NIKP</i>
Log likelihood	33.31	33.57
γ_1	4.485 (0.000)	4.476 (0.000)
ω	1.487 (0.830)	1.558 (0.839)
α_1	0.464 (0.520)	0.484 (0.508)
α_2	-0.188 (0.894)	-0.166 (0.921)
β_1	0.663 (0.630)	0.616 (0.715)

Table 4: Estimation results for the MEM(1,2) model for NIKC and NIKP. P-values for the significance of the coefficients are given in parentheses.

3.3 Diagnostics

Due to the use of the gamma distribution, it is not possible to conduct many standard diagnostic tests for the MMEM models, as such tests assume a normal distribution. Also, as the model has two mixture components and switching between the two regimes is random, there is no straightforward way to obtain residuals. In-sample diagnostic checks can be made by analyzing the so-called probability integral transforms of the data, as proposed by Diebold et al. (1998) and employed by e.g. Bauwens et al. (2004) and Lanne (2006). The probability integral transforms are computed as:

$$z_t = \int_0^{IV_t} f_{t-1}(u) du \quad (6)$$

where $f_{t-1}(\cdot)$ is the conditional density of IV_t relating to the model under analysis. The framework of Diebold et al. (1998) was developed to evaluate density forecasts, but it can be used for in-sample diagnostics as well. The diagnostics are based on the idea that the sequence of probability integral transforms of a model's density forecasts are *iid* uniform $U(0, 1)$ if the model specification is correct. Diebold et al. (1998) recommend the use of graphical procedures to interpret the fit of the models, which makes the approach simple to use and also easily gives clues as to where a misspecification may lie.

Figure 5 plots 25-bin histograms of the probability integral transforms of both NIKC and NIKP with the MMEM(1,2;1,1) model for estimations from the entire in-sample as well as a sample of 500 observations.⁴ All columns fall within the 95% confidence interval based on Pearson's goodness-of-fit test, so the model specification succeeds in taking into account the tails of the conditional distribution for both NIKC and NIKP.⁵ This holds true even when using only 500 observations in the estimation. The Pearson's goodness-of-fit test statistics, as well as the confidence intervals, are not valid as they are

⁴The histograms for the probability integral transforms calculated with the MMEM(1,2;1,2) model are nearly identical.

⁵With a perfect model, z_t would be uniformly distributed and the columns of the histogram would all be of exactly the same height.

calculated without taking estimation error into account. However, this problem most likely leads to rejecting too frequently, and as the p-values from the test range from 0.78 to 0.90, the results lend support to the model specification.

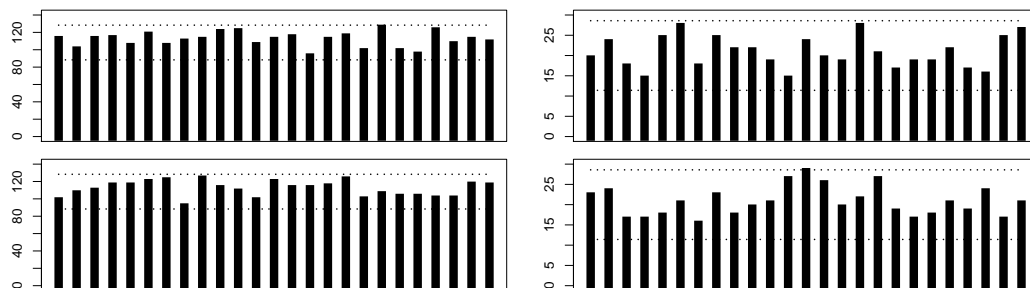


Figure 5: Histograms of probability integral transforms for NIKC (upper left panel) and NIKP (lower left panel) with the MMEM(1,2;1,1) model. Histograms for the MMEM(1,2;1,1) model estimated with the last 500 observations of the in-sample are given on the right. The dotted lines depict the boundaries of the 95% confidence interval.

As a second diagnostic check, autocorrelation functions based on demeaned probability integral transforms and their squares were computed (see Figures 6 and 7).⁶ There is very little autocorrelation in the levels of the demeaned probability integral transforms, but there is noticeable autocorrelation in the squares of demeaned z_t for both NIKC and NIKP. The autocorrelation in squares was also present in the data of Lanne (2006). The situation improves clearly in the sub-sample consisting of only the last 500 observations of the in-sample period.⁷

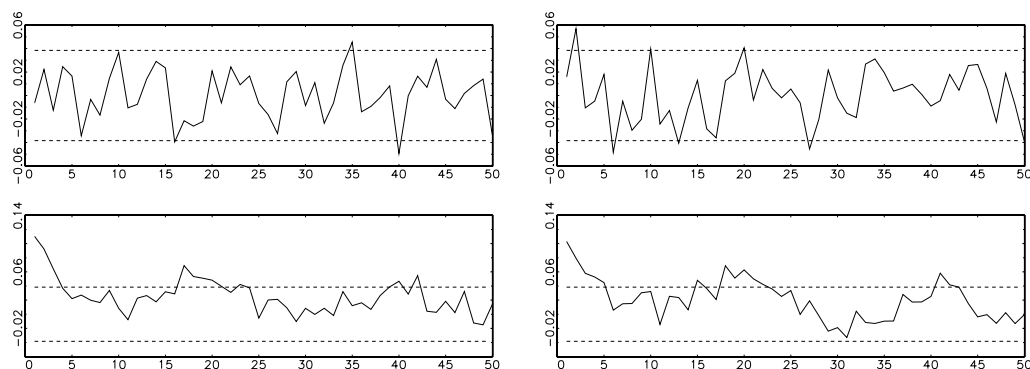


Figure 6: Autocorrelation functions of demeaned probability integral transforms (upper panels) and their squares (lower panels) from the MMEM(1,2;1,1) model. NIKC on left and NIKP on right. The dotted lines depict the boundaries of the 95% confidence interval.

The necessity of using the mixture-MEM model specification is underscored when inspecting the histogram of probability integral transforms calculated with the MEM(1,2) model (Figure 8). With no mixture component, the tails of the conditional distribution

⁶The confidence intervals in the autocorrelation figures are also not valid due to the same estimation error issues as with the Pearson's goodness-of-fit test.

⁷The addition of the statistically insignificant parameter α_{22} to the diagnostic analysis does not improve the autocorrelations.

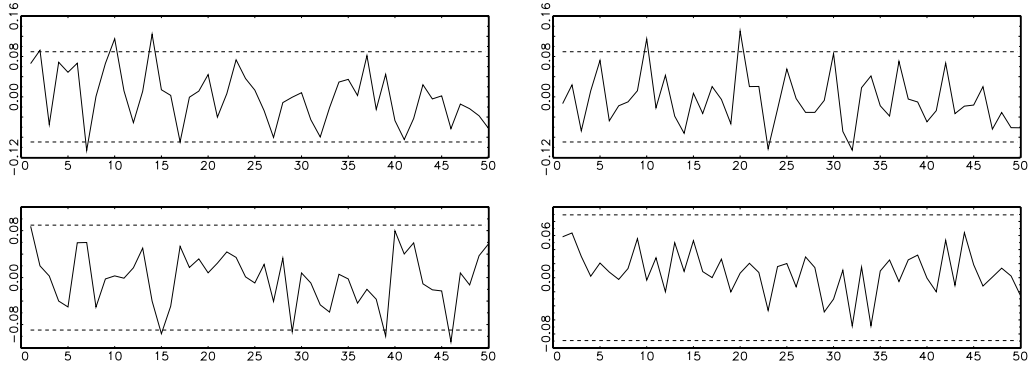


Figure 7: Autocorrelation functions of demeaned probability integral transforms (upper panels) and their squares (lower panels) from the $\text{MMEM}(1,2;1,1)$ model estimated with the last 500 observations of the in-sample. NIKC on left and NIKP on right. The dotted lines depict the boundaries of the 95% confidence interval.

are not modeled properly, with too much emphasis on the mid-range of the distribution. The poor fit of the $\text{MEM}(1,2)$ model is also visible in autocorrelation functions (Figure 9), with autocorrelations from even the level series falling well beyond the confidence interval.

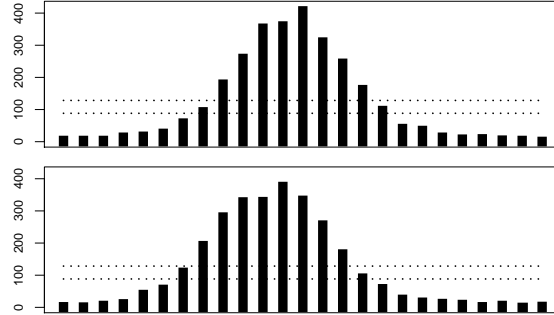


Figure 8: Histograms of probability integral transforms with the $\text{MEM}(1,2)$ model for NIKC (upper panel) and NIKP (lower panel). The dotted lines depict the boundaries of the 95% confidence interval.

4 Forecasts

Forecasts were calculated from the chosen model specification of $\text{MMEM}(1,2;1,1)$ as well as from several competing models in order to assess the value of this modeling approach for option traders and other investors. Of the 3,194 observations in the full sample, the last 486 trading days were left as an out-of-sample period. This corresponds to 1.1.2003 - 31.12.2004.

In addition to the $\text{MMEM}(1,2;1,1)$ model estimated from the entire in-sample, forecasts were calculated from the $\text{MMEM}(1,2;1,1)$ model using 500 observations, as well as from the $\text{MEM}(1,2)$ model, which is expected to fare much worse in the forecast evaluation. Forecasts were calculated by keeping the estimated coefficients constant throughout the out-of-sample period as well as by updating the coefficients each day. In

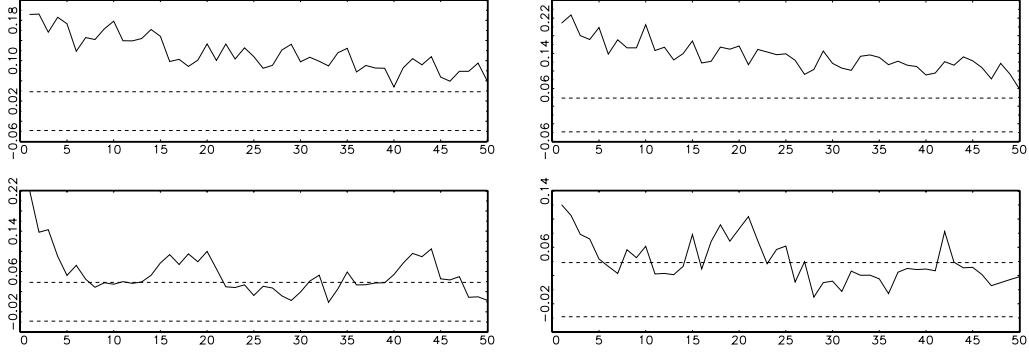


Figure 9: Autocorrelation functions of demeaned probability integral transforms (upper panels) and their squares (lower panels) with the MEM(1,2) model. NIKC on left and NIKP on right. The dotted lines depict the boundaries of the 95% confidence interval.

this case, the sample size was kept constant (2,708 or 500 observations), with the furthest observation dropped and the newest observation added each day. In this alternative, the most recent information is incorporated into the model estimation, which may result in added value if the coefficients are not stable over time.

In practise, the one-step forecasts \widehat{IV}_{t+1} from MMEM models are calculated according to Equation 7:

$$\widehat{IV}_{t+1} = \pi \hat{\mu}_{1,t+1} + (1 - \pi) \hat{\mu}_{2,t+1} \quad (7)$$

For comparison, ARIMA(p,d,q) models were also estimated for the log time series of NIKC and NIKP. This was done in order to see if the MEM specification has added value over more traditional time series models. The chosen specifications are ARIMA(2,0,1) and ARIMA(1,1,1). The former is based on values of the Schwarz Information Criterion for models estimated from the level series and the latter on the previous finding that for the VIX Volatility Index, the ARIMA(1,1,1) specification is best suited for the differenced time series (Ahoniemi (2006)). The ARIMA models

$$IV_t = c_i + \sum_{i=1}^q \alpha_i IV_{t-i} + \sum_{i=1}^p \beta_i \epsilon_{t-i} + \epsilon_t \quad (8)$$

are estimated with rolling samples and coefficients updated each day, similarly to the updating MEM models.

The forecast performance of the various models is summarized in Table 5. Performance is evaluated primarily with directional accuracy, and secondly with mean squared error. Option traders can potentially enter into profitable positions in the market if their expected directional change in IV (up or down) is correct. On the other hand, a forecast of the level of future volatility is of value from a risk management point of view.

In general, it appears to be somewhat easier to forecast NIKC than NIKP, regardless of the model choice. This could perhaps be due to the fact that trading volumes for put options are slightly higher, making put option prices slightly more accurate and predictions more difficult.

The results indicate that the coefficients of the MMEM(1,2;1,1) model are stable over time when using the entire in-sample for estimation. The directional accuracy of

	NIK			NIK _P		
	Correct sign	%	MSE	Correct sign	%	MSE
MMEM(1,2;1,1) - updating	336	69.1%	4.68	321	66.0%	6.02
MMEM(1,2;1,1) - fixed	336	69.1%	4.69	321	66.0%	6.04
MMEM(1,2;1,1) - updating; 500 obs.	325	66.9%	4.59	310	63.8%	5.95
MMEM(1,2;1,1) - fixed; 500 obs.	309	63.6%	4.89	301	61.9%	6.09
MEM(1,2) - updating	314	64.6%	4.93	309	63.6%	6.04
MEM(1,2) - fixed	310	63.8%	4.98	302	62.1%	6.06
ARIMA(2,0,1)	329	67.7%	4.71	321	66.0%	6.06
ARIMA(1,1,1)	327	67.3%	4.71	323	66.5%	6.12

Table 5: Correct sign predictions (out of 486 trading days) and mean squared errors. The best values within each column are in boldface. Mean squared errors for the ARIMA models are calculated by returning log forecasts to the original level.

the model is exactly the same with fixed and updating coefficients for both NIKC and NIKP. Therefore, it would seem that when using a sample period that is sufficiently long, updating brings little benefit and the choice of the sample period is not critical.

The MMEM(1,2;1,1) model forecasts the direction of change correctly on 69.1% of trading days for NIKC and on 66.0% of trading days for NIKP. For e.g. option traders, any level of accuracy over 50% can potentially be worth money. Also, comparing with the findings of Ahoniemi (2006) for the VIX index, whose sign was predicted accurately on 62% of trading days at best, the directional accuracy is clearly better for the Nikkei 225 implied volatility.

When incorporating only the most recent information, or estimating the model with 500 observations, the forecast performance deteriorates considerably. Also, the daily updating of coefficients becomes important, as the directional accuracy improves if using updating rather than fixed coefficients. Further analysis of the parameters estimated with the updating models reveals that a part of the parameter estimates vary considerably over the out-of-sample period. This indicates that 500 observations may be too few to reliably estimate the parameters.

The forecast performance of the MEM(1,2) model falls short of that of the MMEM(1,2;1,1) model estimated from the in-sample, but is no poorer than that of the MMEM(1,2;1,1) model with fixed coefficients estimated from 500 observations. All in all, each model performs at least slightly better in forecasting the direction of change of NIKC rather than NIKP.

When including ARIMA models in the comparison, the MMEM(1,2;1,1) model remains the best predictor for NIKC. However, for NIKP, the ARIMA(1,1,1) model outperforms the MMEM(1,2;1,1) model slightly in simple directional accuracy, and the ARIMA(2,0,1) model is just as accurate. Both the MMEM(1,2;1,1) model and the ARIMA(2,0,1) model produce the correct direction on 321 trading days, with the ARIMA(1,1,1) model generating 323 correct signs. However, the option trading simulation in Section 5 indicates that this difference is most likely not significant - the multiplicative model leads to higher returns.

		Actual outcome		
		<i>Up</i>	<i>Down</i>	<i>Total</i>
Forecast	<i>Up</i>	188	94	282
	<i>Down</i>	56	148	204
	<i>Total</i>	244	242	486

Table 6: NIKC 2×2 contingency table for the MMEM(1,2;1,1) model with updating coefficients

		Actual outcome			
		<i>Up</i>	<i>Zero</i>	<i>Down</i>	<i>Total</i>
Forecast	<i>Up</i>	178	1	101	280
	<i>Down</i>	60	3	143	206
	<i>Total</i>	238	4	244	486

Table 7: NIKP 2×2 contingency table for the MMEM(1,2;1,1) model with updating coefficients

The directional forecasts can be further broken down into contingency tables to investigate where the models make mistakes. Tables 6 and 7 show 2×2 contingency

tables with forecasts from the MMEM(1,2;1,1) model with updating coefficients and actual outcomes. For both NIKC and NIKP, the true number of moves up and down is almost equal, but the model forecasts a move upwards too often. In other words, the model makes more mistakes where the prediction was up but the true change was down than vice versa. There were four days included in the out-of-sample when the change in NIKP was zero, but the change in NIKC non-zero.

For the ARIMA models, the contingency tables are more balanced (see tables 8 and 9 for 2×2 contingency tables for the ARIMA(1,1,1) model). Compared with the contingency tables for the MMEM(1,2;1,1) model, the numbers of up and down forecasts much closer to equal, and therefore, mistakes are made in both directions almost an equal number of times.

		Actual outcome		
		<i>Up</i>	<i>Down</i>	<i>Total</i>
Forecast	<i>Up</i>	161	76	237
	<i>Down</i>	83	166	249
	<i>Total</i>	244	242	486

Table 8: NIKC 2×2 contingency table for the ARIMA(1,1,1) model.

		Actual outcome			
		<i>Up</i>	<i>Zero</i>	<i>Down</i>	<i>Total</i>
Forecast	<i>Up</i>	162	1	83	246
	<i>Down</i>	76	3	161	240
	<i>Total</i>	238	4	244	486

Table 9: NIKP 2×2 contingency table for the ARIMA(1,1,1) model.

The value of the obtained directional forecasts can be assessed with the market timing test developed by Pesaran and Timmermann (1992). The Pesaran-Timmermann test (PT test) stems from the case of an investor who switches between stocks and bonds. The test statistic is computed from contingency tables like the one in Table 6. For NIKP, the days when the actual outcome was 0 are dropped from the analysis in order to run the test. The PT test statistic is computed as in Equation 9 (this version of the test statistic is due to Granger and Pesaran (2000)).

$$PT = \frac{\sqrt{N}KS}{\left(\frac{\hat{\pi}_f(1-\hat{\pi}_f)}{\hat{\pi}_a(1-\hat{\pi}_a)}\right)^{1/2}} \quad (9)$$

where

$$KS = \frac{N_{uu}}{N_{uu}+N_{du}} - \frac{N_{ud}}{N_{ud}+N_{dd}}$$

$$\hat{\pi}_a = \frac{N_{uu}+N_{du}}{N}$$

$$\hat{\pi}_f = \frac{N_{uu} + N_{ud}}{N}$$

N_{uu} is the number of days when both the actual outcome and forecast are up, N_{dd} is the number of days when both the actual outcome and forecast are down, N_{ud} is the number of days when the forecast is up but the actual outcome is down, N_{du} is the number of days when the forecast is down but the actual outcome is up, KS is the Kuiper score, $\hat{\pi}_a$ is the probability that the actual outcome is up, and $\hat{\pi}_f$ is the probability that the outcome is forecast to be up. The limiting distribution of the PT test statistic is $N(0, 1)$ when the null hypothesis that actual outcomes and forecasts are distributed independently is true.

	NIKC		NIKP	
	<i>PT statistic</i>	<i>p-value</i>	<i>PT statistic</i>	<i>p-value</i>
MMEM(1,2;1,1) - updating	8.617	$< 10^{-5}$	7.515	$< 10^{-5}$
MMEM(1,2;1,1) - fixed	8.701	$< 10^{-5}$	7.534	$< 10^{-5}$
MMEM(1,2;1,1) - updating; 500 obs.	7.653	$< 10^{-5}$	6.642	$< 10^{-5}$
MMEM(1,2;1,1) - fixed; 500 obs.	6.722	$< 10^{-5}$	6.028	$< 10^{-5}$
MEM(1,2) - updating	6.898	$< 10^{-5}$	6.451	$< 10^{-5}$
MEM(1,2) - fixed	6.546	$< 10^{-5}$	5.820	$< 10^{-5}$
ARIMA(2,0,1)	7.927	$< 10^{-5}$	7.452	$< 10^{-5}$
ARIMA(1,1,1)	7.748	$< 10^{-5}$	7.614	$< 10^{-5}$

Table 10: Pesaran-Timmermann test statistics and their p-values.

The PT test shows that all the evaluated directional forecast series are statistically significant, as the test statistic has a p-value of less than 0.00001 for all forecast series. In other words, the null hypothesis of predictive failure can be rejected at the one-percent level of significance. The values for the test statistic and their p-values are summarized in Table 10.

When evaluating the mean squared errors of the various forecast series,

$$MSE = \frac{1}{N} \sum_N (\widehat{IV}_t - IV_t)^2 \quad (10)$$

values for NIKC are again superior to those for NIKP (see Table 5). The MMEM(1,2;1,1) model estimated with 500 observations and updating coefficients emerges as the best specification for both NIKC and NIKP. This is perhaps due to the small sample including observations that are relatively near in value to the current level of IV, whereas the entire in-sample contains observations that are tens of percentage points apart.

Mean squared errors can be used in the test for superior predictive ability (SPA) due to Hansen (2005) to check that the forecasts outperform a forecast series of zero change for each day. The SPA test allows for the simultaneous comparison of n series of forecasts, as opposed to the Diebold-Mariano test (Diebold & Mariano, 1995), which compares forecast series one pair at a time. This makes the SPA test useful in the context of this study, as all forecast models can be evaluated at the same time. In the SPA test, one series of forecasts is defined to be the benchmark, and MSE is then used as a loss function to determine whether that benchmark is the best forecast series. The null hypothesis of the test is that the benchmark is not inferior to the other forecast series.

Two benchmarks are used for the purposes of this paper. The first is zero change, meaning that today's value for NIKC or NIKP is used as the forecast for the next day. The consistent p-value of the test, calculated using 1,000 bootstrap resamples, is reported as less than 0.00001 for both NIKC and NIKP. Therefore, there is at least one other forecast series that is a better forecaster than the benchmark of zero change. In other words, the forecasts calculated in this study can clearly add value to such a naive forecast.

The second benchmark is the MMEM(1,2;1,1) model estimated with the entire in-sample and updating coefficients, which produces the second-lowest MSE for both time series. Although being only second best, this model yields much better directional accuracy for both NIKC and NIKP than the model with the lowest MSE. The SPA test reports consistent p-values of 0.448 for NIKC and 0.675 for NIKP with this benchmark, meaning that there is little evidence against the null hypothesis that the benchmark is not inferior to the other series of forecasts.⁸

Longer-term forecasts

Investors seeking volatility forecasts in order to evaluate the risks in their portfolios may be more interested in longer-term forecasts than simply one day ahead. To address this issue, forecasts up to five trading days ahead in time have been calculated for both NIKC and NIKP from the models presented above. Table 11 summarizes the mean squared errors for these forecasts.

Model	MSE (NIKC)	MSE (NIKP)
MMEM(1,2;1,1) - updating	6.94	10.46
MMEM(1,2;1,1) - fixed	7.25	12.16
MMEM(1,2;1,1) - updating; 500 obs.	8.65	10.77
MMEM(1,2;1,1) - fixed; 500 obs.	59.46	12.55
MMEM(1,2) - updating	8.90	10.70
MMEM(1,2) - fixed	9.18	10.71
ARIMA(2,0,1)	7.19	10.68
ARIMA(1,1,1)	7.98	12.05

Table 11: Mean squared errors from forecasts five trading days ahead. Best results for each column in boldface. Mean squared errors for the ARIMA models are calculated by returning log forecasts to the original level.

The longer horizon advocates the use of the mixture multiplicative model with updating coefficients. Not only is the mean squared error for the MMEM(1,2;1,1) model with updating coefficients the lowest for both NIKC and NIKP, there is clear value in updating coefficients each day to include the most recent information. The performance of the ARIMA models is quite good as well, but nevertheless falls short of the values received from the best MMEM model. The poor performance of the MMEM(1,2;1,1) model with 500 observations and fixed coefficients for NIKC reflects the fact that a part of the parameter estimates change greatly during the out-of-sample period if updated.

⁸The consistent p-values of the SPA test when using the third model as the benchmark, or the model with the lowest MSE, are 0.907 and 0.909 for NIKC and NIKP, respectively.

The SPA test confirms the statistical significance of the ranking, giving a consistent p-value of 0.940 to the MMEM(1,2;1,1) model with the lowest mean squared error for NIKC and 0.908 for NIKP.

5 Option trading

In order to further assess the value of the forecasts calculated in Section 4, an option trading simulation was carried out with market quotes for Nikkei 225 index options. If an option trader can correctly forecast the direction of change in implied volatility over the next trading day, with all other factors held equal, it is possible to enter into profitable positions in options markets. The purpose of this simulation is not to determine whether trading based on these forecasts could lead to abnormal returns, but to use the returns to rank the forecast models.

The straddle is an option spread that is particularly useful for traders with a view on the coming movements in IV. Straddles are volatility trades where a long position benefits from a rise in IV, and a short position benefits from a fall. Noh et al. (1994) use straddles in their option trading simulation, and Bollen & Whaley (2004) note that the most effective way to trade given an impending upward move in volatility is to buy straddles. In a long straddle, an equal number of call and put options are purchased, with a short position involving selling an equal number of calls and puts.

Option quotes for Nikkei 225 index options were obtained for the entire out-of-sample period, or 1.1.2003 - 31.12.2004. Straddle returns were calculated based on the directional forecasts of all the forecast series reported above. Positions were entered into at the closing quote of each day, and closed with the closing quote of the following day. A theoretical sum of 1,000 Japanese Yen is invested in long positions, and 1,000 Yen is received from short positions, so that the absolute value of the options does not affect the returns. Harvey and Whaley (1992) and Noh et al. (1994) also use such a fixed investment outlay.

The data on Nikkei 225 options that was used in this analysis contains strike prices 500 index points apart. On each day, the strike price that was as close-to-the-money as possible was selected. Trading volumes are usually highest for at-the-money options and, as noted by Bollen & Whaley (2004), at-the-money options are most sensitive to volatility. The straddle positions are not delta-hedged. At-the-money straddles are close to delta-neutral, and in this particular study, the emphasis is on comparing the forecast series, not determining the best possible exact outcome of the option trading. Due to this consideration, no transaction costs are deducted from the returns. Options with the next maturity date were used up until two calendar weeks before expiration, with trading then rolled on to the next maturity date. This switch ensures that any odd movements in option quotes and volatility close to expiration do not affect the results.

In practise, the return from a long straddle (R_l) was calculated as in Equation 11, and the return from a short straddle (R_s) as in Equation 12. The proceeds from the short position are held with zero interest.

$$R_l = \frac{1000}{C_{t-1} + P_{t-1}}(-C_{t-1} - P_{t-1} + C_t + P_t) \quad (11)$$

$$R_s = \frac{1000}{C_{t-1} + P_{t-1}}(C_{t-1} + P_{t-1} - C_t - P_t) \quad (12)$$

C_{t-1} is the close quote of the near-the-money call option that is used to enter into the straddle, P_{t-1} is the respective quote for the put option with the same strike price and maturity date, and C_t and P_t are the closing quotes of the same options at the end of the next trading day.

It may not be beneficial to trade every single day, particularly if the signal for change in IV is very weak. Three filters have been used to leave out the forecasts for the smallest changes in IV, i.e. treat a part of the forecasts as zero-change forecasts. This approach is similar to those used by Harvey and Whaley (1992) and Noh et al. (1994), who each used two filters in their trading simulations. Poon and Pope (2000) use three filters to account for transaction costs. The three filters employed in this study leave out forecasts for a change of under 1.0%, under 2.0%, and under 5.0%, respectively. As these filters are too severe for the ARIMA(1,1,1) model, leading to less than ten trades with the 5.0% filter, filters of 0.1%, 0.5%, and 1.0% are used in this one case.

The option trading returns are presented in Table 12. For NIKC, the forecasts from the MMEM(1,2;1,1) model estimated with the full in-sample yield the highest returns. When selecting a forecast model from the MEM family, the use of a filter would not seem necessary. The evidence for ARIMA models is the opposite, with filtering improving the returns for both models.

When looking at the returns from NIKP forecasts, the highest value is again received with the MMEM(1,2;1,1) model, but with the use of fixed coefficients. As with NIKC, the use of a filter decreases the returns for the MMEM(1,2;1,1) model when it is estimated using the full in-sample. Interestingly, the use of a filter is profitable for half of the MEM models (those with weakest directional forecast ability). This lends further support to using the MMEM(1,2;1,1) specification with a long in-sample as the forecast model of choice.

As with NIKC, all returns from ARIMA models for NIKP fall short of the best return with a MMEM model. This occurs despite the fact that the directional accuracy of the ARIMA(2,0,1) model was equal to that of the MMEM(1,2;1,1) model, and the ARIMA(1,1,1) model produced the correct sign on two days more than the MMEM(1,2;1,1) model. Consistent with the directional forecast accuracy that was presented in Table 5, the MMEM(1,2;1,1) model with updating coefficients leads to positive option trading returns on 51.9% of the out-of-sample trading days, the MMEM(1,2;1,1) model with fixed coefficients and the ARIMA(2,0,1) model lead to positive returns on 52.3% of the trading days, and the ARIMA(1,1,1) model has a positive outcome on 52.7% of the trading days.⁹ However, one particular outlier that negatively affects the returns from the ARIMA(1,1,1) model is the return for trading with the closing prices of 27.2.2004 and 1.3.2004. NIKP rose by 17% during that period, but the signal from the ARIMA(1,1,1) model was for a move downwards. This resulted in a loss of 1,152 Yen on that day.

⁹Even with perfect foresight, i.e. by entering into straddles knowing what direction the IV will take over the coming day, the trading return would be positive only on 61.3% of the days for NIKP in this particular out-of-sample. This reflects the fact that many other market parameters that affect option prices also change continually (in addition to volatility).

		NIK						NIK _P						
		Trading P/L		Filter I	Filter II	Filter III	Trading P/L	Filter I	Filter II	Filter III	Trading P/L	Filter I	Filter II	Filter III
MMEM(1,2;1,1) - updating MMEM(1,2;1,1) - fixed MMEM(1,2;1,1) - updating; 500 obs. MMEM(1,2;1,1) - fixed; 500 obs.		21769	19442 (353)	18944 (288)	11138 (120)		19797	16374 (386)	17264 (296)	14613 (113)		16374 (386)	17264 (296)	14613 (113)
		21369	19225 (359)	19050 (288)	11101 (115)		19953	16025 (380)	17138 (301)	13808 (109)		16025 (380)	17138 (301)	13808 (109)
		21202	20604 (393)	18025 (312)	14836 (174)		18417	17870 (414)	17402 (348)	16844 (178)		17870 (414)	17402 (348)	16844 (178)
		20274	19563 (419)	18508 (347)	15422 (185)		16188	18269 (416)	18982 (355)	16620 (188)		18269 (416)	18982 (355)	16620 (188)
		18868	17655 (418)	17983 (358)	14920 (200)		16739	19114 (407)	16506 (343)	16087 (182)		19114 (407)	16506 (343)	16087 (182)
MEM(1,2) - updating MEM(1,2) - fixed ARIMA(2,0,1) ARIMA(1,1,1)		18463	17477 (417)	18075 (360)	14877 (203)		17575	18433 (404)	17973 (346)	15971 (184)		18433 (404)	17973 (346)	15971 (184)
		18677	20494 (464)	20610 (447)	20077 (388)		18968	18088 (472)	17501 (458)	17355 (412)		18088 (472)	17501 (458)	17355 (412)
		20993	21124 (461)	19306 (322)	15342 (213)		16598	16169 (455)	18568 (331)	13894 (200)		16169 (455)	18568 (331)	13894 (200)

Table 12: Option trading returns from the out-of-sample period of 486 trading days. The return from trading on all 486 days is given in columns Trading P/L. The best values for NIK and NIK_P are given in boldface. Filter I gives no trading signal when the projected change in IV is under 1.0 %, Filter II when the signal is under 2.0 %, and Filter III when the signal is under 5.0 % (with the exception of the ARIMA(1,1,1) model, for which Filter I = 0.1 %, Filter II = 0.5 %, and Filter III = 1.0 %). The number of days traded out of 486 is given in parentheses for each filtered return.

The SPA test was employed in order to assess the statistical significance of the differences in the unfiltered option trading returns. In this application, the opposite numbers of the returns achieved when trading every day were used as the loss function in the test. For NIKC, the logical benchmark is the first forecast model. The consistent p-value in this case is 0.992, indicating that there is little evidence against the null hypothesis that this is the best forecast model. For NIKP, the same model, or MMEM(1,2;1,1) with updating coefficients and the full in-sample, was used as the benchmark model despite its being only second best in option trading returns. The use of this benchmark is supported by its consistent p-value of 0.800.¹⁰ Therefore, the evidence from the option trading simulation as a whole is not in conflict with the choice of the MMEM(1,2;1,1) model with updating coefficients.

6 Conclusions

A multiplicative error model with two alternative mean equations and two alternative gamma distributions for the error term was estimated for time series of implied volatilities derived from call and put options on the Nikkei 225 index. The mixture-MEM model was found to be a good fit, possessing statistically significant coefficients and satisfactory in-sample diagnostics. Without the mixture components, the model is a much worse fit to the data.

Measured with directional accuracy, one-step-ahead forecasts calculated from various MMEM models outperform those from ARIMA models for the time series of call IV in a two-year out-of-sample period. An ARIMA model fares marginally better than MMEM models for the put-side implied volatility. Again, the lack of a mixture component leads to poorer forecasts. When mean squared errors are used for forecast evaluation, MMEM models are superior for both call and put IV.

Further analysis lends support to recommending the use of a MMEM model for forecasting the implied volatility of the Nikkei 225 index. MMEM models produce lower mean squared errors when forecasting five trading days ahead in time, and an option trading simulation that calculates returns from long and short straddle positions based on the signals of the forecasts indicates that the highest returns can be achieved when trading based on a MMEM model. Statistical tests confirm this superiority.

These results indicate that option traders and others interested in forecasting the direction of change of implied volatility in the Japanese market can benefit from using the new class of multiplicative models, as directional accuracy is well over 50 percent. The directional forecasts are also statistically significant. A mixture specification seems to be necessary in order to obtain the best possible results. Investors looking to forecast the future level of volatility implied by Nikkei 225 options or the future level of volatility in the returns of the Nikkei 225 index can also receive added value from the forecasts of MMEM models, at least up to five trading days into the future. Trading in the Nikkei 225 index option market can also be shown to result in positive returns before transaction costs if entering into straddle positions based on the forecasts from MMEM models - at least in the particular out-of-sample used in this study.

¹⁰The consistent p-value when using the MMEM(1,2;1,1) model with fixed coefficients as the benchmark is 0.999.

REFERENCES

- Ahoniemi, K. (2006), 'Modeling and Forecasting Implied Volatility - an Econometric Analysis of the VIX Index', *HECER Discussion Paper No. 129*
- Bauwens, L. & Giot, P. (2003), 'Asymmetric ACD models: Introducing price information in ACD models', *Empirical Economics*, 28, 709-731
- Bauwens, L., Giot, P., Grammig, J., & Veredas, D. (2004), 'A comparison of financial duration models via density forecasts', *International Journal of Forecasting*, 20, 589-609
- Blair, B.J., Poon, S-H., & Taylor, S.J. (2001), 'Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns', *Journal of Econometrics*, 105, 5-26
- Bollen, N.P.B. & Whaley, R.E. (2004), 'Does Net Buying Pressure Affect the Shape of Implied Volatility Functions?', *The Journal of Finance*, 59, 711-753
- Brooks, C. & Oozer, M.C. (2002), 'Modeling the Implied Volatility of Options on Long Gilt Futures', *Journal of Business Finance & Accounting*, 29, 111-137
- Canina, L., & Figlewski, S. (1993) 'The Informational Content of Implied Volatility', *The Review of Financial Studies*, 6, 659-681
- Christensen, B.J., & Prabhala, N.R. (1998), 'The relation between implied and realized volatility', *Journal of Financial Economics*, 50, 125-150
- Day, T.E. & Lewis, C.M. (1992), 'Stock market volatility and the information content of stock index options', *Journal of Econometrics*, 52, 267-287
- Diebold, F.X., Gunther, T.A., & Tay, A.S. (1998), 'Evaluating Density Forecasts with Applications to Financial Risk Management', *International Economic Review*, 39, 863-883
- Diebold, F.X. & Mariano, R.S. (1995), 'Comparing Predictive Accuracy', *Journal of Business & Economic Statistics*, 13, 253-263
- Engle, R.F. (2002), 'New Frontiers for ARCH models', *Journal of Applied Econometrics*, 17, 425-446
- Engle, R.F. & Gallo, G.M. (2006), 'A multiple indicators model for volatility using intra-daily data', *Journal of Econometrics*, 131, 3-27
- Engle, R.F. & Russell, J.R. (1998), 'Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data', *Econometrica*, 66, 1127-1162
- Fernandes, M. & Grammig, J. (2006), 'A family of autoregressive conditional duration models', *Journal of Econometrics*, 130, 1-23
- Ghysels, E., Gouriéroux, C. & Jasiak, J. (2004), 'Stochastic volatility duration models', *Journal of Econometrics*, 119, 413-433
- Granger, C.W.J & Pesaran, M.H. (2000), 'Economic and Statistical Measures of Forecast Accuracy', *Journal of Forecasting*, 19, 537-560

- Hansen, P.R. (2005), 'A Test for Superior Predictive Ability', *Journal of Business & Economic Statistics*, 23, 365-380
- Harvey, C.R. & Whaley, R.E. (1992), 'Market volatility prediction and the efficiency of the S&P 100 index option market', *Journal of Financial Economics*, 31, 43-73
- Jorion, P. (1995), 'Predicting Volatility in the Foreign Exchange Market', *The Journal of Finance*, 50, 507-528
- Lanne, M. (2006), 'A Mixture Multiplicative Error Model for Realized Volatility', *Journal of Financial Econometrics*, 4, 594-616
- Manganelli, S. (2005), 'Duration, volume, and volatility impact of trades', *Journal of Financial Markets*, 8, 377-399
- Meitz, M. & Teräsvirta, T. (2006), 'Evaluating Models of Autoregressive Conditional Duration', *Journal of Business & Economic Statistics*, 24, 104-124
- Nelson, D.B. & Cao, C.Q. (1992), 'Inequality Constraints in the Univariate GARCH Model', *Journal of Business & Economic Statistics*, 10, 229-235
- Noh, J., Engle, R.F. & Kane, A. (1994), 'Forecasting Volatility and Option Prices of the S&P 500 Index', *The Journal of Derivatives*, 2, 17-30
- Pesaran, M.H. & Timmermann, A.G. (1992), 'A simple non-parametric test of predictive performance', *Journal of Business & Economic Statistics*, 10, 461-465
- Poon, S-H., & Pope, P.F. (2000), 'Trading volatility spreads: a test of index option market efficiency', *European Financial Management*, 6, 235-260